

Clustering Methods for Statistical Downscaling in Short-Range Weather Forecasts

J. M. GUTIÉRREZ AND A. S. COFIÑO

Department of Applied Mathematics, E.T.S.I. Caminos, University of Cantabria, Santander, Spain

R. CANO

Instituto Nacioal de Meteorología, CMT/CAS, Santander, Spain

M. A. RODRÍGUEZ

Instituto de Física de Cantabria, CSIC, Santander, Spain

(Manuscript received 18 March 2003, in final form 13 October 2003)

ABSTRACT

In this paper an application of clustering algorithms for statistical downscaling in short-range weather forecasts is presented. The advantages of this technique compared with standard nearest-neighbors analog methods are described both in terms of computational efficiency and forecast skill. Some validation results of daily precipitation and maximum wind speed operative downscaling (lead time 1–5 days) on a network of 100 stations in the Iberian Peninsula are reported for the period 1998–99. These results indicate that the weighting clustering method introduced in this paper clearly outperforms standard analog techniques for infrequent, or extreme, events (precipitation > 20 mm; wind > 80 km h $^{-1}$). Outputs of an operative circulation model on different local-area or large-scale grids are considered to characterize the atmospheric circulation patterns, and the skill of both alternatives is compared.

1. Introduction

During the last two decades the skill of numerical atmospheric circulation models (ACMs) used for short- and medium-range weather prediction have increased substantially because of the advances both in assimilation procedures and physical parameterizations. Current ACMs integrated by different weather services simulate accurately the atmospheric synoptic dynamics on coarse-grained 40–100-km-resolution grids. However, at finer spatial resolution these models have much smaller skill, since the physical parameterization of subgrid-scale processes—such as cloud formation, evaporation, orography, turbulence, etc.—is a difficult task and, moreover, the parameters may not be tuned for a particular region of interest.

Several methods have been proposed in the literature with the aim of gaining subgrid detail in these predictions; these techniques are referred to as downscaling methods (see, e.g., Wilby and Wigley 1997; Zorita and von Storch 1999). On the one hand, the so-called dynamic downscaling methods use the ACM-integrated

gridded fields as boundary conditions for a new higher-resolution limited-area model (LAM) that includes parameterizations adapted for the region of interest—typically a few hundred kilometers. However, systematic errors from the ACM could also transmit to LAMs and, therefore, some filtering postprocessing technique is needed to eliminate such trends (Kalman filters; Bergman and Delleur 1985a,b, among others).

On the other hand, the availability of historic climate data led to the development of statistical techniques. These statistical downscaling methods work with climatological databases of observations (e.g., precipitation, wind speed, and temperature) from a representative number of stations (gauges or sites) within the area of study. These observations are statistically related to the gridded atmospheric patterns, leading to forecast models for adapting a gridded forecast to local climates in a straightforward way. For instance, given a database of atmospheric circulation patterns \mathbf{x}_i , used as predictors, and simultaneous historical records of a local variable y_i (predictand), standard statistical methods such as regression analysis can be applied to obtain a linear model $\hat{y}_i = \mathbf{a}\mathbf{x}_i + b$; some applications are described in Enke and Spekat (1997) and Billet et al. (1997). Other techniques such as model output statistics (Klein and Glahn 1974) or canonical correlation analysis (Barnett and

Corresponding author address: Prof. José Manuel Gutiérrez, Departamento de Matemática Aplicada, E.T.S.I. Caminos, Universidad de Cantabria, 39005 Santander, Spain.
E-mail: gutierjm@unican.es

Preisendorfer 1987) follow a similar scheme and also suffer from the same limitation: they assume normality and linear dependence among the variables. These assumptions are accomplished when dealing with seasonal anomalies but, in short-range weather forecasts, variables such as precipitation require a more general framework. These limitations can be addressed considering nonlinear models $\hat{y}_i = f(\mathbf{x}_i)$, where the functional form of f is fitted using modern nonparametric techniques (feed-forward neural networks; Gardner and Dorling 1998; McGinnis 1994, among others). However, these global models assume stationary atmosphere dynamics during the period of available data (note that a single global model is trained with the available data), and this is by no means guaranteed. Local techniques such as the method of analogs (nearest neighbors) provide a simple solution to this problem, since instead of training a global model to data, different models are trained for different local regions of the reanalysis database; thus, these models are naturally adapted to the trends and variations of the system dynamics.

The local method of analogs introduced by Lorenz (1969) in the framework of time series prediction is a particular implementation of the more general nearest-neighbors (NN) methodology. This method assumes that similar circulation flows \mathbf{x}_i and \mathbf{x}_j lead to similar local outcomes y_i and y_j . Therefore, a local prediction \hat{y}_i can be derived from an ensemble of analogs of a forecast of the atmospheric flow $\hat{\mathbf{x}}_i$ provided by an ACM. This ensemble is formed by those historical days for which the observed atmospheric flow (characterized by the analysis of a reanalysis database) is closest to the forecast. In other words, this method provides a local framework to train downscaling models using the information of the reanalysis database most similar to the low-resolution gridded forecast. Some implementations of this method for detecting climatic anomalies (see Zorita and von Storch 1999; Wilby and Wigley 1997; and references therein) and for short-range forecast (see, e.g., Van den Dool 1989) have been presented in the literature. In general, it has been shown that the analog method performs as well as other complicated downscaling techniques (see, e.g., Zorita and von Storch 1999) indicating that these poor-man methods are efficient alternatives for many downscaling problems.

The different applications of the analog or nearest-neighbors methodology differ basically in the following characteristics:

- The specific atmospheric variables and the large-scale or local-area grid selected for characterizing the atmospheric flow. The resulting vectors \mathbf{x}_i are referred to as atmospheric patterns. There has been some controversy about the convenience of considering large- or local-scale area grids for defining the atmospheric patterns (Van den Dool 1989). Over a small area it is easy to find good analogs but, in this case, boundary and remote effects can travel to the target grid point

in the course of the prediction period, thus deteriorating the downscaled forecast. On the other hand, large-scale atmospheric patterns are more robust for characterizing the atmospheric state for a 12- or 24-h prediction, but some patterns may have no close analog in the database to make a skillful forecast. In this paper we compare the performance of both choices for downscaling precipitation and maximum wind speed on the Iberian Peninsula.

- The algorithm used to obtain the ensemble of analogs $\{\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_k}\}$ for \mathbf{x}_i from the database of atmospheric patterns (the subindices t_1, \dots, t_k refer to the dates of the analog days). The NN technique with a prescribed ensemble size k (k -NN) is the most simple and common method used for this task. However, this algorithm is computationally intensive, since for each forecast pattern the set of analogs is computed from the whole reanalysis database. In this paper we compare different clustering (or quantization) techniques that reduce significantly the computational cost, since the reanalysis database is partitioned into a reduced number of groups (clusters); each of these groups is represented by a reference, or prototype, pattern. Thus, each cluster can be considered a natural “ensemble of analogs” for the patterns within the group. This downscaling algorithm is computationally efficient, since the only task required for an operative run is selecting the representative cluster for a given forecast pattern. The main shortcoming of the clustering technique is the reduction of the variance and the resolution of the forecast, due to the quantization process, when compared with standard k -NN. To overcome this limitation we present a weighting method and show how the resulting technique outperforms k -NN in terms of computational cost and skill for extreme events.
- The method used to infer a forecast \hat{y}_i for a given local variable y_i (predictand), based on the ensemble of analogs. Different methods have been proposed for this task, including linear regression and neural networks for numeric predictions. In the case of probabilistic forecasts, the empirical distribution given by the ensemble of analogs $\{y_{t_1}, \dots, y_{t_k}\}$ is usually considered. However, it is well known that this estimation is bumpy, especially when the empirical distribution is drawn from a small set (i.e., small k , or a small cluster). Several alternatives have been proposed in the literature to avoid this problem. In this paper we use a simple technique of kernel density estimation to obtain the probability function of the local variable based on the ensemble of analogs.

In this paper we introduce a new clustering-based statistical downscaling method considering the above-mentioned characteristics and compare its performance with the standard analog algorithm. Different grids are used to define the atmospheric patterns to be used for downscaling, and several experiments are conducted

considering daily precipitation and maximum wind speed in a network of 98 climatic stations on the Iberian Peninsula from 1997 to 1998. The best results are obtained considering a local-scale grid (including a temporal component) and using a weighted version of the clustering downscaling method.

In section 2 we describe the area of study and the data used in this work; we also analyze the problem of dimensionality reduction using principal component analysis. Both a limited-area and a large-scale grid are considered as different alternatives for characterizing atmospheric circulations patterns. Section 3 gives a brief introduction to clustering algorithms. The proposed downscaling method is presented in section 4, where we also describe some validation experiments in detail. Finally, some conclusions and further remarks are given in section 5.

2. Data and dimensionality reduction

One of the main requisites for performing statistical downscaling using analog techniques is maintaining the consistency between the operative and the reanalysis circulation models. The operative ACM is used to obtain a predicted gridded atmospheric flow and the reanalysis database is searched to obtain a similar historical days (to get the ensemble of analogs). In this work we use the analysis given by the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis Project 15 (ERA-15; <http://www.ecmwf.int/research/era/>), obtained by integrating a T106L31 model for the period from 1979 to 1993. This provides us with a set of fields with daily temperature (T), relative humidity (H), geopotential (Z) and U , V wind components at six pressure levels (300, 500, 700, 850, 925, and 1000 mb) at 0000, 0600, 1200, 1800, and 2400 UTC (other variables and levels were also stored, but we have tried to make the dimension of the problem tractable). The vector comprising the resulting information for a given day and grid area is called the atmospheric pattern. The reanalysis model is a simpler version of the operative ECMWF T511L60, from which we take the forecast patterns to feed the downscaling methods. Therefore, the consistency among the models is maintained in this work.

a. Local data of meteorological stations

Regarding the local climate records database, we compiled daily precipitation (precip) and maximum wind speed (wind) from the primary network of 98 climatic stations provided by the Spanish National Weather Service-Instituto Nacional de Meteorología (INM; see Fig. 1). These stations cover different hydrographic regions, as shown in Fig. 1b, and thus spatial considerations have to be taken into account when validating the methods. These variables are measured daily at 0600 UTC. Thus the 24-h forecast period ranges from 0600 to 0600 UTC the following day. In all cases, daily data

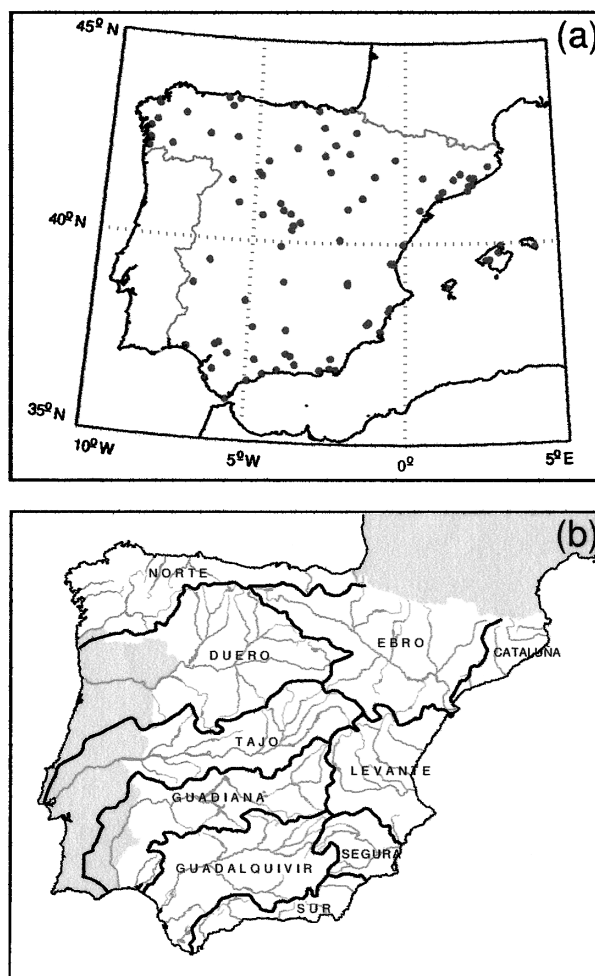


FIG. 1. (a) Primary network of 98 climatic stations in the Iberian Peninsula and Balearic Islands. (b) Main hydrographic basins for the Iberian Peninsula.

are available from 1979 to present, containing missing values in different proportions. An advantage of downscaling methods based on analogs is that missing values can be handled very easily, since missing observations can be discarded from the ensemble of analogs. The 98 stations considered in this work have been selected to guarantee that they contain at most 10% of missing observations during the period of analysis (used as training data) and validation (test).

Probabilistic forecasts for a variable y (precipitation or maximum wind speed) are obtained in terms of probabilities $P(y > u)$, where u is an appropriate threshold for y . In the case of precipitation, we considered the thresholds precip > 0.1 , 10, and 20 mm; for maximum wind speed we analyzed the thresholds wind > 50 and 80 km h^{-1} .

b. Atmospheric patterns

The geographical area of interest in this work is the Iberian Peninsula. With the aim of comparing the effi-

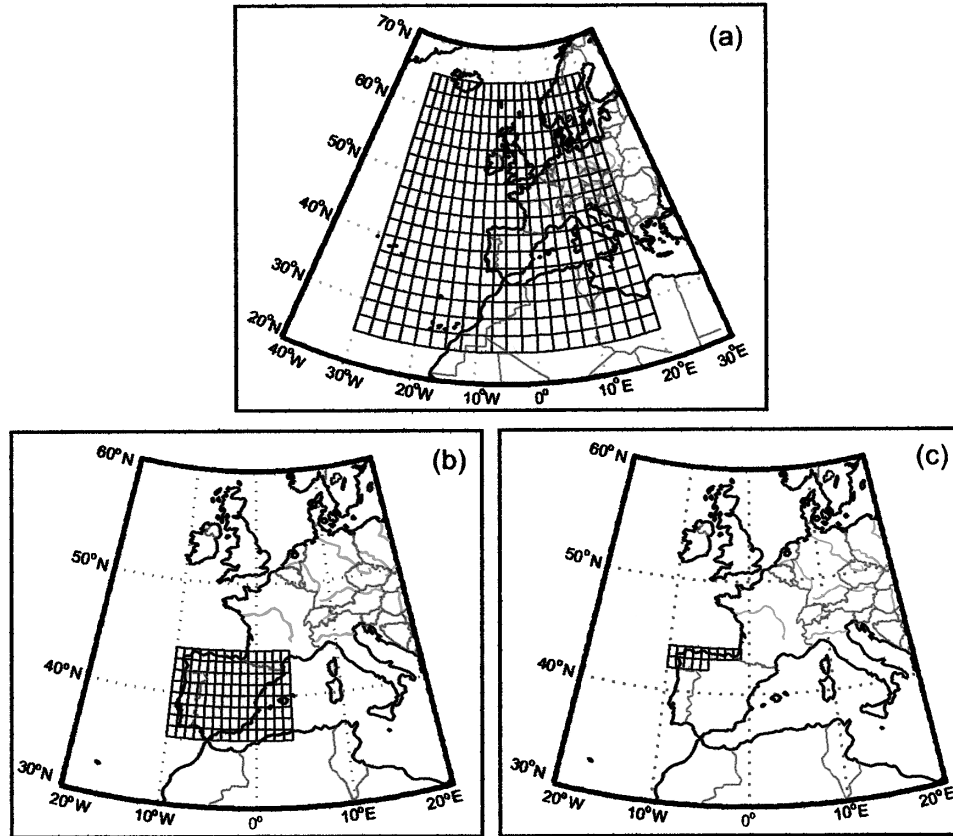


FIG. 2. Maps of the model grid domains used in this study: (a) large-scale macro- β grid considered for model 1, (b) meso- α grid covering the peninsula for model 2, and (c) meso- β model 3 grid for the northern basin. (Twelve different grids were considered, one for each basin of the Iberian Peninsula. For the sake of clarity only the north basin is shown.)

ciency of different grid areas and temporal resolutions, we restrict the reanalysis and operative fields to a 3D macro- β grid centered in the Iberian Peninsula (model 1), a 4D meso- α grid covering the peninsula (model 2), and a 4D meso- β grid constrained to a basin of interest (model 3)—in this case, we consider a specific grid for each of the 12 basins in the Iberian Peninsula. The resulting atmospheric patterns become

- model 1: The $2.5^\circ \text{ lat} \times 2.5^\circ \text{ lon}$ grid shown in Fig. 2a. In this case, the patterns are obtained by combining the T , H , Z , U , and V fields at 1200 UTC at the six mentioned pressure levels:

$$\mathbf{x}_{12} = (T_{12}^{1000}, \dots, T_{12}^{300}, H_{12}^{1000}, \dots, H_{12}^{300}, \dots, V_{12}^{1000}, \dots, V_{12}^{300}), \quad (1)$$

where X_i^j denotes the j pressure level field of variable X variable at i UTC.

- model 2: The $1.0^\circ \times 1.0^\circ$ grid shown in Fig. 2b covering the area of study. In this case, we use the same fields T , H , Z , U , and V , but we include a temporal component by taking the fields both at 0600 on the forecast day and 24-h later, referred to here as \mathbf{x}_{30} (note that observations correspond to the same peri-

od). This temporal component of the pattern compensates the scale reduction of the grid, since it accounts for the boundary and remote effects that may come into the local grid during the forecast period:

$$\mathbf{x} = (\mathbf{x}_{06}, \mathbf{x}_{30}). \quad (2)$$

- model 3: The $1.0^\circ \times 1.0^\circ$ grid shown in Fig. 2c. In this case, we consider a specific pattern for each of the 12 basins in the Iberian Peninsula, combining the same fields on a higher-resolution temporal domain: 0600, 1200, 1800, 2400, and 0600 UTC. In this case, we cover the forecast period with all the available temporal information:

$$\mathbf{x} = (\mathbf{x}_{06}, \mathbf{x}_{12}, \mathbf{x}_{18}, \mathbf{x}_{24}, \mathbf{x}_{30}). \quad (3)$$

Note that the data considered in all models are very highly dimensional. For instance, we get a total of 20×20 (grid) $\times 6$ (pressure levels) $\times 5$ (variables) = 12 000 dimensions for characterizing each daily atmospheric circulation pattern with model 1.

c. Dimensionality reduction

Although the atmospheric patterns are highly dimensional, they are highly correlated, both among different

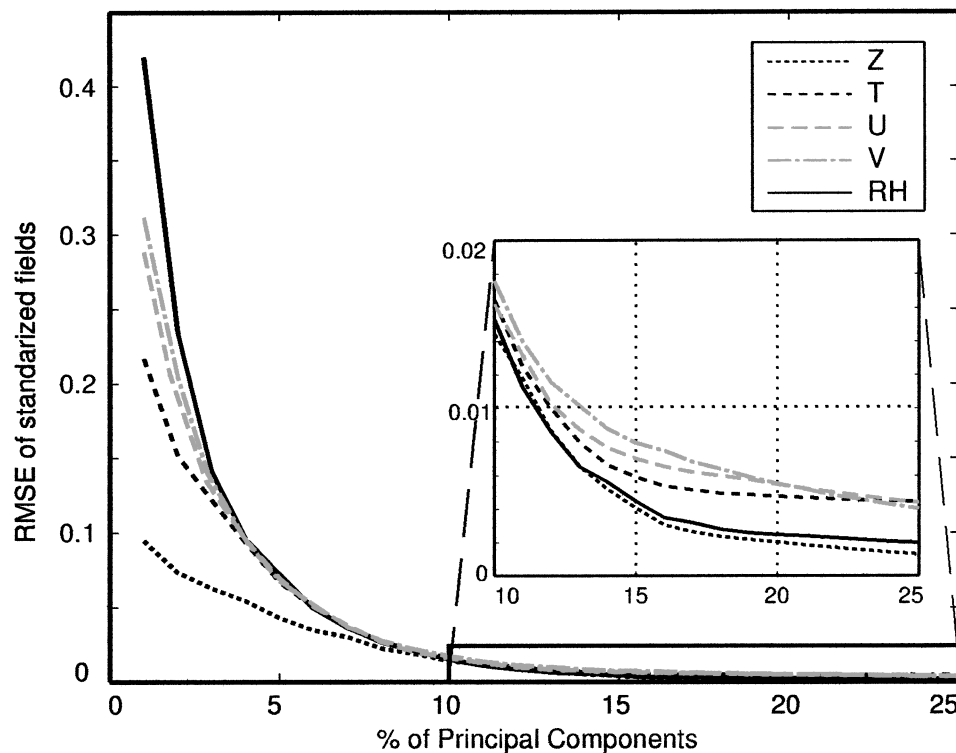


FIG. 3. Reconstruction rmse for each of the five variables in model 1 (the error is computed for the 3D standardized fields) considering an increasing number of PCs (ranging from 1% of the original vector dimension to 25%).

horizontal and vertical levels of a given variable and among different variables. This problem slows down (and even spoils) the training process of clustering algorithms. Therefore, we have previously considered the problem of dimensionality reduction, to extract as much correlation as possible from the patterns but keeping their discriminant power. Principal component analysis (PCA) is a classical statistical linear method that has been widely used in data analysis and compression since it gives an optimal (in the mean-square sense) linear reduction of dimension [see Preisendorfer and Mobley (1988) for an overview of PCA in meteorology]. For instance, given the $15 \times 365 = 5475$ ERA-15 reanalysis atmospheric 12 000-dimensional patterns \mathbf{x}_i , as described in (1), the basic idea of PCA is to find d linearly transformed components \mathbf{w}_j of the original vectors, so that the transformed d -dimensional vectors $\mathbf{u}_i = \mathbf{x}_i M$ give the smaller reconstruction error $\sum_i [\mathbf{x}_i - (M^T \mathbf{u}_i^T)^T]^2$ when projected back to the original space, where M is the $12\,000 \times d$ orthogonal base transform matrix with j th column given by \mathbf{w}_j . To avoid problems due to different scales, all the variables are previously standardized for each grid point and pressure level before applying PCA.

In order to select a convenient dimension reduction for each of the alternative models described above, we have computed the reconstruction root-mean-square error (rmse) of the patterns for an increasing number of

principal components (PCs) (ranging from 1% to 25% of the original dimension) for each of the models. For instance, Fig. 3 shows the results for model 1, where the reconstruction error of each of the five variables is computed separately. Note that, although the PCs are obtained globally, combining all the variables in the pattern vector, the reconstruction errors are similar for all of them. Only in the case of considering a low number of components is the reconstruction significantly more efficient with those variables with smoother fields (e.g., Z or T).

From Fig. 3 we can see that 10% of the original dimension leads to a reconstruction error lower than 2% of the standard deviation of each of the 3D variable fields. Therefore, we can reduce the dimension of the atmospheric patterns from 12 000 to ≈ 1000 with no significant information loss (assimilation errors are usually higher than the above reconstruction error). Note that an alternative criterion for selecting an appropriate number of PCs could be based on the average distance between the patterns corresponding to analog days in the reanalysis database. The number of PCs could be lowered, keeping only the most significant linear modes with no alteration of the results, but in this paper we have used the reconstruction error as an objective criterion for dimensionality reduction (further research is needed to analyze the role of each of the PCs in the

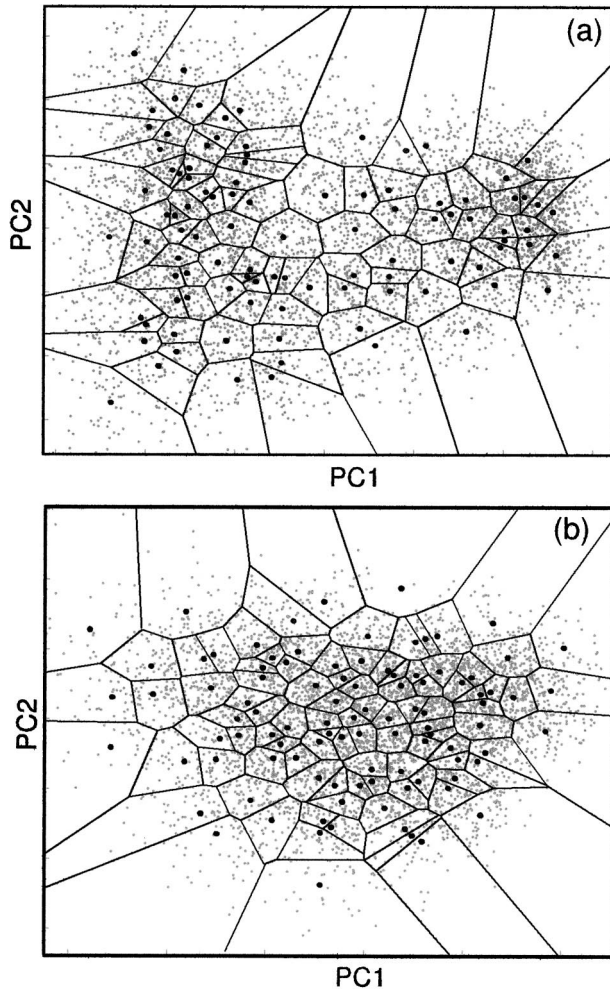


FIG. 4. Clustering of the ERA-15 reanalysis database with the m -means algorithm considering 100 clusters for (a) model 1 and (b) model 3. The graph shows the daily patterns and centroids projected onto the space spanned by the two leading principal components. The separation lines between different clusters are also shown (these lines correspond to the Voronoi diagram associated with the centroids).

downscaling process, but this problem is beyond the scope of this paper).

The same analysis was also performed for models 2 and 3, obtaining the PCs needed for an operative application of the algorithm.

3. Clustering techniques

An important limitation of the standard k -NN method is the equal number of members (reanalysis daily patterns) considered in all the ensembles of analogs. It is well known that the circulation patterns do not follow a multinormal distribution—they only follow such a distribution in a radial sense for a certain season (Toth 1991). Therefore, considering ensembles of analogs of a fixed size k is not consistent with the distribution of the reanalysis database patterns (some regions are much

more sparse than others, and ensembles of variable size are required). The clustering approach is a simple solution for this problem, since cluster sizes are automatically adapted to the inhomogeneities of the distribution.

The most convenient clustering methods for a large number of patterns in high-dimensional space are the iterative centroid adjustment algorithms [see Duda et al. (2000) for a detailed introduction to classification and clustering techniques]. The most common of these methods is the m -means algorithm. Given a database $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, of d -dimensional real vectors [e.g., atmospheric patterns in (1)–(3)], and a prescribed number of groups m , the m -means algorithm computes a set of d -dimensional prototypes or centroids $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$, each of them characterizing a group of data $\mathbf{C}_i \subset \mathbf{X}$ formed by the vectors in the database for which \mathbf{v}_i is the nearest prototype. This task is accomplished following an iterative procedure, which starts from an initial set of centroids $\mathbf{v}_1^0, \dots, \mathbf{v}_m^0$, chosen at random [see Peña et al. (1999) for a description and comparison of different initialization procedures]. The goal of the algorithm is to minimize an overall within-cluster distance from the patterns to the centroids:

$$\sum_{i=1, \dots, m} \sum_{\mathbf{x}_j \in \mathbf{C}_i} \|\mathbf{x}_j - \mathbf{v}_i\|^2. \quad (4)$$

Since an exhaustive minimum search is prohibitive, a local minimum is computed by iteratively adjusting the cluster centroids, and by reassigning each pattern to the closest centroid. On the $(r + 1)$ th iteration, each of the j th vectors \mathbf{x}_j is assigned to the i th group, where $i = \text{argmin}_i \|\mathbf{x}_j - \mathbf{v}_i^r\|$ [$\text{argmin}_c f(c)$ gives the value c with minimum $f(c)$], and the prototypes are updated as the mean of the corresponding patterns:

$$\mathbf{v}_i^{r+1} = \sum_{\mathbf{x}_j \in \mathbf{C}_i} \mathbf{x}_j / \#\mathbf{C}_i, \quad (5)$$

where $\#\mathbf{C}_i$ denotes the number of elements in \mathbf{C}_i . Under certain conditions, the above iterative processes converge after R iterations, and the final centers \mathbf{v}_i^R are the prototypes (centroids). Each of the centroids \mathbf{v}_i represents a cluster \mathbf{C}_i that consists of the patterns closer to \mathbf{v}_i^R than to any other centroid. Thus, the m -means clustering algorithm consists of the following steps:

- 1) Select the number of desired clusters m .
- 2) Initialize the cluster centers (e.g., randomly).
- 3) Repeat.
 - (a) Assign each vector (atmospheric pattern) to its closest cluster center;
 - (b) Recompute the centers for each cluster, to be the mean of the patterns assigned to that cluster.

The above algorithm was applied to the 5500 reanalysis patterns of ERA-15, characterizing the atmospheric configuration for a period of 15 yr. We considered different numbers of clusters $m = 100, 200$, and 400 , which correspond to different mean numbers of elements in

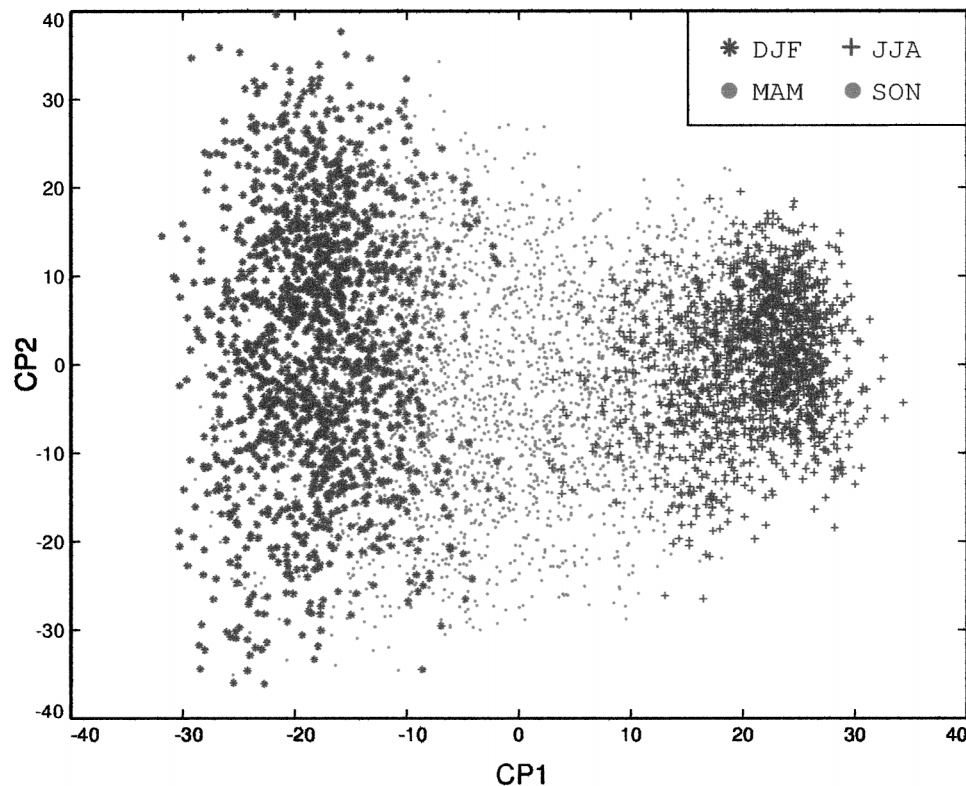


FIG. 5. Seasonal ERA-15 daily patterns corresponding to model 1 projected onto the space spanned by the two leading principal components. The bimodality of the distribution is given by the first PC and corresponds to the annual cycle. The legend shows the symbols used for each of the seasons.

each cluster: approximately 50, 25, and 15, respectively; we shall use this relation later to compare this method with standard k -NN. For instance, considering $m = 100$ we obtained the prototypes shown in Figs. 4a and 4b for the models 1 and 3, respectively. Note that the distribution of daily patterns corresponding to model 1 is clearly bimodal in the space spanned by the first two leading PCs. This fact is given by the first PC, which is associated with an empirical orthogonal function (EOF) corresponding to the different seasonal variations of the atmospheric flow in different latitudes on the large-scale grid (see Fig. 5). The local model 3 distribution is not affected by this phenomena, since the seasonal variations are similar in the range of latitudes corresponding to this grid.

4. Clustering-based downscaling method

Different downscaling implementations of the method of analogs have been presented in the literature, each including a specific algorithm to obtain the ensemble of analogs for a gridded forecast pattern (the low-resolution input of the downscaling process). The standard application of this methodology uses the k -nearest neighbors of the forecast pattern in the reanalysis database as the ensemble of analogs (see, e.g., Zorita and von Storch 1999). This process is time consuming since

it involves the calculation of distances from the forecast pattern to all the patterns in the reanalysis database. In this paper we present an alternative method based on the clustering techniques described in section 3. This algorithm allows us to reduce the computation time by partitioning the database into meaningful subgroups (weather classes). Each of the resulting groups C_i —characterized by a prototype \mathbf{v}_i —is used as the ensemble of analogs for those patterns assigned to the cluster, that is, those patterns closer to \mathbf{v}_i than any other prototype. The clustering process also deals with the inhomogeneities of the reanalysis distribution, since the number of elements in each of the clusters is automatically adapted according to the distribution of atmospheric patterns.

Figure 6 shows the schemes of the standard k -NN and clustering-based analog downscaling techniques. As shown in Fig. 6b, the main advantage of the clustering downscaling technique is that the reanalysis data are replaced by a predetermined number of clusters C_1, \dots, C_m with associated prototypes $\mathbf{v}_1, \dots, \mathbf{v}_m$, respectively (afterward, the reanalysis database is no longer needed). This is done applying the m -means method as a preprocessing step (as done with principal components, which are not shown in the figure for the sake of clarity). In the operative phase, the application of both the k -NN analog and clustering techniques to obtain the ensemble

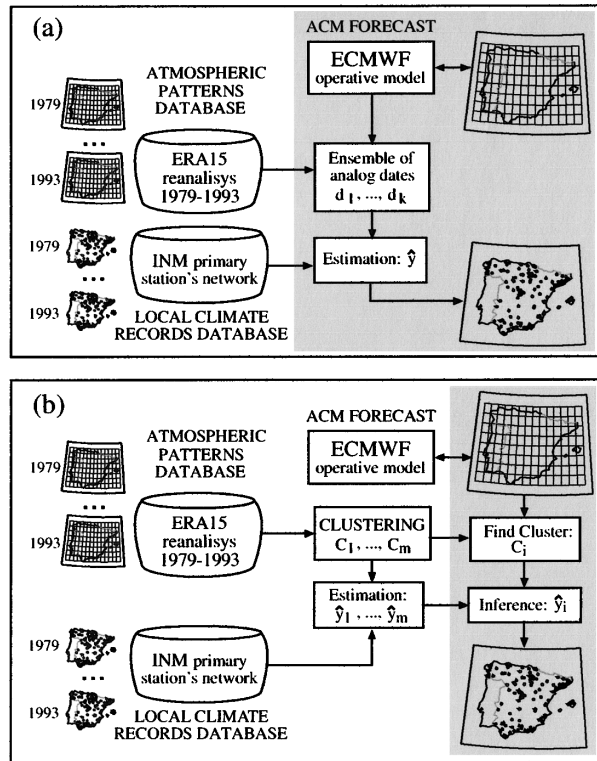


FIG. 6. (a) Scheme of the standard k -NN analog downscaling technique, and (b) scheme of the clustering downscaling algorithm. In both cases, a database of atmospheric patterns and a database of local climate records is required. The methods start from an initial gridded forecast (the forecast pattern) and end up with predictions for the available local stations. The shaded areas indicate the operative components of the algorithms required to obtain a local forecast.

of analogs for a given forecast pattern \mathbf{x}_t is based on a brute force calculation of distances. However, in the case of the analog method, distances from \mathbf{x}_t to each pattern in the database are required, whereas in the clustering method the only distances needed are between \mathbf{x}_t and \mathbf{v}_i , $i = 1, \dots, m$. Note that this reduces dramatically the computational load of the algorithm.

In the case of the k -NN analog technique, given a given forecast pattern \mathbf{x}_t , the set of analogs $\{\mathbf{x}_{t_1}, \dots,$

$\mathbf{x}_{t_k}\}$ is obtained (note that t denotes the current date of the forecast, whereas t_1, \dots, t_k refer to dates corresponding to the reanalysis period), and a local forecast \hat{y}_t^s for a station s and a variable y (precipitation or maximum wind speed) is obtained from the set of past observations $\{y_{t_1}^s, \dots, y_{t_k}^s\}$.

On the other hand, in the case of the clustering downscaling method we can compute a local forecast for each cluster \mathbf{C}_i from the observations $\{y_{t_1}^s, \dots, y_{t_{q(i)}}^s\}$ corresponding to the dates of the reanalysis patterns contained in the cluster $\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_{q(i)}}$, where $q(i)$ denotes the number of elements of cluster \mathbf{C}_i . Thus, the posterior probability of the local variable y^s for cluster i can be estimated from the empirical distribution:

$$P_i^s = P(y^s > u | \mathbf{C}_i) = \frac{\#[y_{ij}^s > u; j = 1, \dots, q(i)]}{q(i)}. \quad (6)$$

Then, in the operative application of the method, if a forecast pattern \mathbf{x}_t is assigned to the cluster \mathbf{C}_i , the local forecast \hat{y}_t^s will be directly obtained as P_i^s in (6). However, it is well known that this estimation is bumpy, especially for small $q(i)$ (which may be the case for some clusters resulting from the clustering process). Several alternatives have been proposed to overcome this limitation, including the simple technique of kernel density estimation (Hastie et al. 2001). In this case, each observation is considered not as a single real number, but as a kernel function (usually a Gaussian kernel) centered at the given point:

$$\phi_\lambda(x, x_0) = \frac{1}{2\pi\lambda^2} \exp\left(-\frac{(x - x_0)^2}{2\lambda}\right). \quad (7)$$

The kernel process smooths the estimation by adding independent Gaussian noise (the kernels) to the observations. This acts like a filter convolved with the empirical distribution. In this case, an estimation of the probability is given by the integral of the density function:

$$f(y^s) = \frac{1}{q(i)} \sum_{j=1}^{q(i)} \phi_\lambda(y^s, y_{ij}^s). \quad (8)$$

TABLE 1. Annual spatial averaged RSA for precipitation using Models 1, 2, and 3 for lead-time forecasts ranging from 1 to 3 days ($D + 1$, $D + 2$, and $D + 3$). Results for the three different methods are reported: Analog (k -NN with $k = 50$), Cluster (m -means with $m = 100$), and WCluster weighted (m -means with $m = 400$ and $w = 4$).

| Fore-cast | Method | >0.1 mm | | | >10.0 mm | | | >20.0 mm | | |
|-----------|----------|---------|-------|--------------|----------|-------|--------------|----------|-------|--------------|
| | | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| D + 1 | Analog | 0.647 | 0.750 | <u>0.791</u> | 0.602 | 0.728 | <u>0.776</u> | 0.480 | 0.643 | 0.773 |
| | Cluster | 0.538 | 0.682 | 0.744 | 0.501 | 0.627 | 0.710 | 0.427 | 0.583 | 0.681 |
| | WCluster | 0.597 | 0.733 | <u>0.783</u> | 0.574 | 0.715 | <u>0.769</u> | 0.526 | 0.685 | <u>0.781</u> |
| D + 2 | Analog | 0.633 | 0.737 | <u>0.771</u> | 0.586 | 0.689 | <u>0.727</u> | 0.474 | 0.606 | 0.647 |
| | Cluster | 0.523 | 0.669 | 0.716 | 0.478 | 0.602 | 0.667 | 0.408 | 0.535 | 0.614 |
| | WCluster | 0.588 | 0.711 | <u>0.763</u> | 0.571 | 0.685 | <u>0.736</u> | 0.504 | 0.647 | <u>0.729</u> |
| D + 3 | Analog | 0.572 | 0.693 | <u>0.734</u> | 0.572 | 0.674 | <u>0.694</u> | 0.467 | 0.602 | 0.624 |
| | Cluster | 0.449 | 0.640 | 0.678 | 0.449 | 0.576 | 0.631 | 0.372 | 0.513 | 0.591 |
| | WCluster | 0.542 | 0.680 | <u>0.726</u> | 0.542 | 0.668 | <u>0.706</u> | 0.489 | 0.632 | <u>0.675</u> |

TABLE 2. Annual spatial averaged RSA for maximum wind speed using Models 1, 2, and 3 for lead-time forecasts ranging from 1 to 3 days.

| Fore- cast | Method | >50 km h ⁻¹ | | | >80 km h ⁻¹ | | |
|---------------|----------|------------------------|-------|--------------|------------------------|-------|--------------|
| | | 1 | 2 | 3 | 1 | 2 | 3 |
| D + 1 | Analog | 0.576 | 0.702 | <u>0.721</u> | 0.500 | 0.556 | 0.511 |
| | Cluster | 0.453 | 0.609 | <u>0.648</u> | 0.428 | 0.511 | 0.512 |
| | WCluster | 0.526 | 0.671 | <u>0.716</u> | 0.524 | 0.670 | <u>0.697</u> |
| D + 2 | Analog | 0.574 | 0.682 | <u>0.707</u> | 0.491 | 0.598 | <u>0.590</u> |
| | Cluster | 0.421 | 0.583 | <u>0.630</u> | 0.384 | 0.521 | 0.549 |
| | WCluster | 0.514 | 0.653 | <u>0.708</u> | 0.472 | 0.703 | <u>0.706</u> |
| D + 3 | Analog | 0.562 | 0.657 | <u>0.668</u> | 0.476 | 0.552 | <u>0.572</u> |
| | Cluster | 0.428 | 0.567 | <u>0.605</u> | 0.359 | 0.497 | 0.548 |
| | WCluster | 0.508 | 0.630 | <u>0.656</u> | 0.468 | 0.652 | <u>0.620</u> |

From this equation we can obtain different statistics of the distribution needed for the forecast process. In our case, we shall obtain an estimation of (6) for the different threshold values of the variable.

Thus, the clustering downscaling algorithm is computationally efficient, since the only task required for an operative run is selecting the representative cluster for a given forecast pattern and returning the forecast associated with the cluster (see the shaded area in Fig. 6b).

However, the main shortcoming of the clustering technique is the reduction of the variance and the resolution of the forecast, due to the quantization process and the associated border effects. Note that the same forecast is given for all daily patterns within a cluster, independently of their relative position in the cluster (close to the center or to the boundary). Thus, only m different forecast values are given by the algorithm for a given station. To overcome this limitation we consider a weighted version of the algorithm given by

$$\hat{y}_t^s = \frac{\sum_{c=1}^w d(c)P_{t(c)}^s}{\sum_{c=1}^w d(c)}, \quad (9)$$

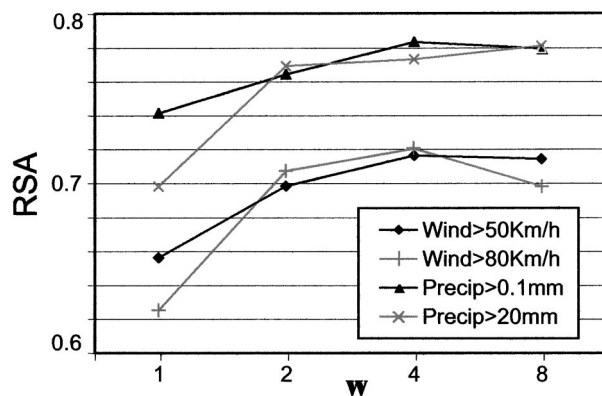


FIG. 7. RSA for precipitation (events precip > 0.1 mm and precip > 20 mm) and maximum wind speed (events wind > 50 km h⁻¹ and wind > 80 km h⁻¹) using the weighting m -means method in Eq. (9) with w ranging from 1 to 8 and $m = 100 \times w$ (to keep the averaged number of elements close to 50).

where $t(c)$ is the index of the c th closest cluster to \mathbf{x}_t , $d(c)$ is the distance from $\mathbf{v}_t(c)$ to \mathbf{x}_t , and $P_{t(c)}^s$ is the estimation of the probability obtained for cluster $t(c)$ using (8). Instead of assigning the same estimation for all patterns within a cluster, a weighted sum of the estimations associated with the clusters closer to each particular pattern is considered. This modification also overcomes the problem of resolution reduction, since now the possible forecasts are not limited to m cases. We refer to this method as the weighted clustering downscaling method.

In this paper we are only interested in the predictive skill of the proposed methodology when compared with the standard analog algorithm and in the analysis of different grid pattern specifications (models 1, 2, and 3). Therefore, we do not focus on the problem of correcting the predicted probability function according to the observed climatology, or some other “inflation” methodology, so that the predicted variance matches the observed variance [the reader is referred to von Storch (1999), and references therein, for more details about this problem].

Verification and comparison

In order to verify the skill of the weighted clustering downscaling method (WCluster) to forecast daily precipitation (precip) and maximum wind speed (wind), we use the standard k -NN analog method (Analog) as a benchmark. We compare different values of the parameters for these algorithms. On the one hand, the Analog method only depends on the number of analogs k considered to make a forecast. On the other hand, the WCluster method depends on the number of clusters considered (m), and the number of neighboring clusters to be weighted in the prediction (w). Note that the reanalysis database contains approximately 5000 data patterns and then, similar results are expected for values k and m such that $5000/m = k$. Moreover, to avoid a reduction of the variance due to the smoothing associated with the weighting process, the ratio m/w is kept constant, so an increment of the parameter w will lead to a similar increment in the number of clusters.

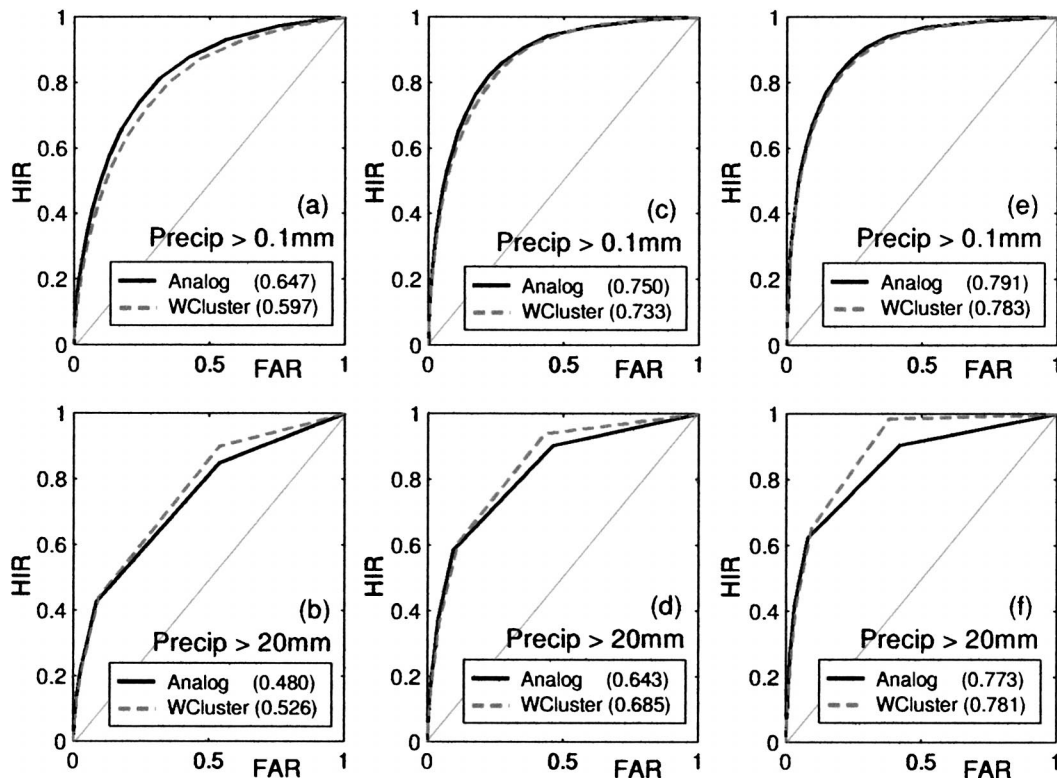


FIG. 8. ROC curves and RSA (in parentheses) for precipitation [events precip > 0.1 mm and precip > 20 mm] using the k -NN analog method with $k = 50$ (label Analog), and weighted m -means with $m = 400$ and $w = 4$ (label WCluster) for models (a)–(b) 1, (c)–(d) 2, and (e)–(f) 3.

We have considered operative data (lead time 1 to 5 days) from the period 1998–99 with no overlap using ERA-15. Predictions for each station in Fig. 2 are treated as individual forecasts, and the skills of the different stations are averaged into the final outcome. Some of the results are detailed for each of the four seasonal forecast periods: fall [September–October–November (SON)], winter [December–January–February (DJF)], spring [March–April–May (MAM)], summer [June–July–August (JJA)], and others are reported as annual averages (1998–99). We consider different binary events given by the thresholds precip > 0.1 , 10, and 20 mm, and wind > 50 and 80 km h^{-1} , as shown in section 2a. A forecast is given in terms of the probability estimated for the occurrence of the event [e.g., $p(\text{event}) = p(\text{precip} > 0.1 \text{ mm})$] using the nearest-neighbors (or cluster) ensemble of analogs. Relative operating characteristic (ROC) curves are a standard verification methodology for this type of probabilistic forecast [see Katz and Murphy (1997) for a detailed description]. The calculation of an ROC curve is based on a contingency table given by the number of observed occurrences and nonoccurrences of an event versus the simultaneous forecasted occurrences and nonoccurrences of the event; to this aim, the event is forecasted to occur when the estimated probability reaches a threshold [$p(\text{event}) \geq q$]; then for each different probability threshold q , a different con-

tingency table is obtained. Two parameters are used to characterize each of these contingency tables: HIR_q (hit rate) measures the rate of successful predictions [$p(\text{predicted}|\text{occurred})$], and FAR_q (false alarm rate) measures the rate of wrong predictions [$p(\text{predicted}|\text{not occurred})$]. The ROC curve is obtained plotting HIR_q versus FAR_q for the different values of q in $[0, 1]$. The area A under the ROC curve gives the skill of the forecast method—the ROC skill area index (RSA) is obtained as $2A - 1$ and varies between zero (no skill) and one (perfect forecast). The RSA score has several interesting advantages when compared with other standard scores such as Brier skill score (BSS). For instance, BSS is insensitive to infrequent (rare, or extreme) events, whereas the RSA takes account of this fact (HIR is conditioned to the occurred events, whereas FAR is conditioned to the not-occurred events).

In order to check the performance of the downscaling method considering the three pattern areas used in this paper (see Fig. 1), we considered four different values for k (100, 50, 25, and 15) for the k -NN method, and four related m values of the number of clusters (50, 100, 200, and 400) for the m -means algorithm. The best results—shown in Tables 1 and 2—were obtained with model 3 with an intermediate number of elements $k = 50$ (label Analog) and $m = 100$ for the m -means algorithm (label Cluster). We also considered different

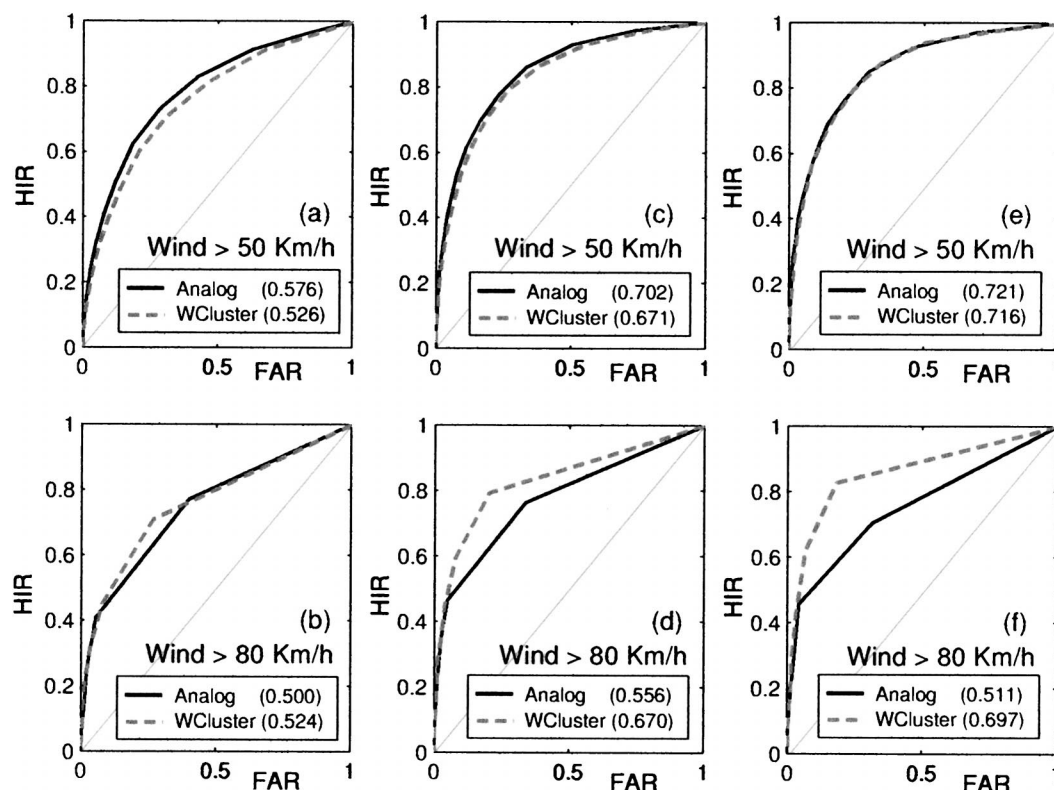


FIG. 9. As in Fig. 8, but ROC curves and RSA for maximum wind speed (events wind $> 50 \text{ km h}^{-1}$ and wind $> 80 \text{ km h}^{-1}$).

values of parameters w and m of the weighted clustering method (label WCluster). The performance of the method as a function of the weighting factor w is shown in Fig. 7, which displays the evolution of the RSA for different events when considering different combination of the parameters; in all cases, the ratio m/w is kept close to 100, the optimum number of clusters obtained. In this case, a convenient choice of the values is shown to be given by the configuration $m = 400$, $w = 4$ for the WCluster algorithm.

Table 1 shows the annual spatial averaged RSA precip values for the 98 stations shown in Fig. 2, and Table 2 reports the results for the case of wind for 1–3-days lead time. Similarly, Figs. 8 and 9 compare the ROC curves and RSA scores corresponding to the optimal configurations of parameters obtained for the Analog kNN and WCluster algorithms for 1-day lead time ($D + 1$). From these tables and figures we can conclude the following:

- 1) As expected, the analog method clearly outperforms the basic clustering downscaling algorithm.
- 2) The weighted clustering downscaling method outperforms the analog technique for high precipitation and wind thresholds (extreme events: precip $> 20 \text{ mm}$; wind $> 80 \text{ km h}^{-1}$). For nonextreme events both methods present similar skills. Figure 10 shows the scatterplots of RSA values versus the climatological frequency of four different events for the 98

climatic stations. This figure shows the relationship between the skill of the forecasts and the frequency of observation of the event in a particular station. In the case of extreme events, we can see from Figs. 10b and 10d that the skill tends to be lower in those stations where the frequency of the event is smaller (the most rare the event, the less skilful the prediction is likely to be). However, we can see that, in this situation, the skill of the weighted clustering downscaling method (labeled WCl in Fig. 10) is larger than the skill of the standard Analog algorithm.

- 3) The best results are obtained with model 3, indicating that the optimal definition for the atmospheric flow is a 4D pattern restricted to the local geographical domain of interest.
- 4) The skill decays as the lead time of the forecast used to feed the downscaling methods increases from 1 to 3 days. Figure 11 shows this fact in more detail, considering lead times from 1 to 6 days. This figure shows that the decline is more pronounced from lead time of 4 days, indicating a horizon for the skill of short-range weather forecast ACMS.

Note that the above results were based on spatial and temporal averaged skills, which may hide some important aspects of the forecasts for different regions of the Iberian Peninsula and for different seasons. For this reason, some extra validations were performed to analyze

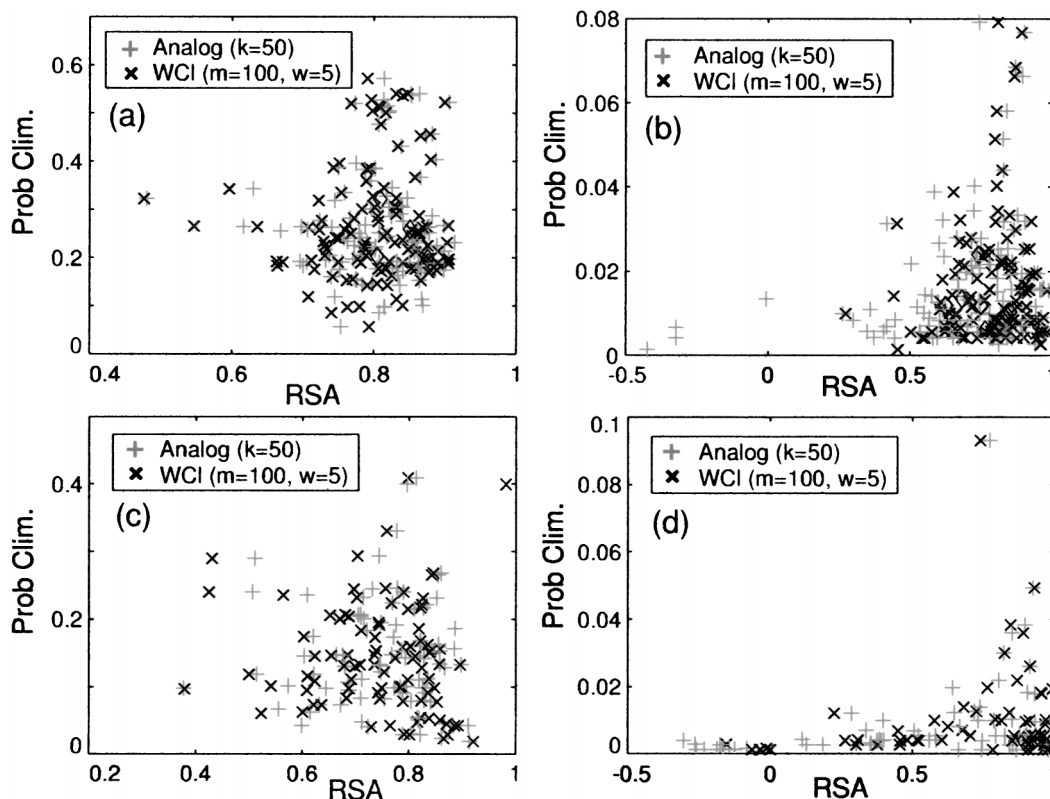


FIG. 10. Scatterplots of RSA values vs the climatological frequency of the event for the 98 climatic stations: (a) precip > 0.1 mm, (b) precip > 20 mm, (c) wind < 50 km h^{-1} , and (d) wind < 80 km h^{-1} , for the weighted clustering downscaling (WCluster) and the k -NN (Analog) algorithm.

the temporal and spatial distribution of skill. In particular, Table 3 shows the RSA values for the 12 hydrological basins in the Iberian Peninsula (see Fig. 2b). Annual and seasonal averages are given for each of the optimal models found in the previous analysis: Analog method with $k = 50$, and WCluster with $m = 400$ and

$w = 4$. From this table we can see how the weighted clustering algorithm outperforms the standard analog method for precip > 20 mm, whereas both methods are comparable for nonextreme events. Among the different seasons, summer exhibits the lowest skills in most of the basins. The reason for this could be the dominance of convective precipitation during the summer season in the Iberian Peninsula. Downscaling low-scale convective precipitation from atmospheric circulation patterns is more difficult than downscaling large-scale precipitation associated with fronts. This fact will be investigated further in a future work.

Finally, Fig. 12 shows some details about the daily errors of the method to forecast precipitation. To this aim, a period of 90 days was considered (September, October, and November 1998), and the daily errors [Brier scores (BS)] for the 98 stations were computed using the weighted clustering (WCluster with $m = 400$ and $w = 4$) and standard analog method (k -NN with $k = 50$). The differences between these errors were characterized using a box-and-whisker plot (the boxes represent the terciles for the 98 stations, and the whiskers show the extreme differences that occurred each day). Negative values indicate stations where the clustering method outperforms the analog technique. From this figure we can see that there is a large variability of the

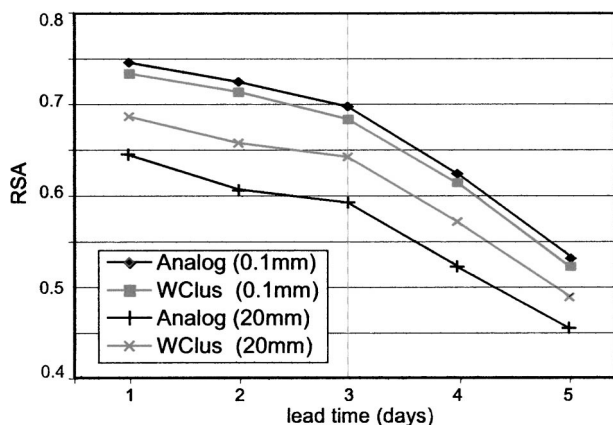


FIG. 11. RSA vs the lead time of the forecast for precipitation (events precip > 0.1 mm and precip > 20 mm) using the analog k -NN method (with $k = 50$) and the weighting m -means method (with $m = 400$ and $w = 4$). A pronounced decay is shown in all cases for a lead time of 4 days (model 2 has been used in this figure).

TABLE 3. Regional averaged seasonal RSA for precipitation using the k -NN analog method with $k = 50$ (label k NN), and weighted m -means with $m = 400$ and $w = 4$ (label WCI). The climatic frequency of each of the events is also shown (label Clm). The symbol “—” indicates that no validation data are available for the period. The column on the left shows the 12 hydrological regions in the Iberian Peninsula.

| Region | Method | >0.1 mm | | | | >20 mm | | | |
|--------------|--------|---------|-------|-------|-------|--------|-------|-------|-------|
| | | DJF | MAM | JJA | SON | DJF | MAM | JJA | SON |
| Segura | k NN | 0.831 | 0.779 | 0.415 | 0.808 | 0.832 | 0.818 | — | 0.783 |
| | WCI | 0.836 | 0.800 | 0.476 | 0.831 | 0.930 | 0.977 | — | 0.829 |
| | Clm | 0.223 | 0.169 | 0.035 | 0.190 | 0.013 | 0.006 | 0.000 | 0.006 |
| Balears | k NN | 0.717 | 0.804 | 0.600 | 0.726 | 0.625 | 0.826 | — | 0.662 |
| | WCI | 0.722 | 0.794 | 0.624 | 0.719 | 0.750 | 0.831 | — | 0.774 |
| | Clm | 0.270 | 0.195 | 0.067 | 0.330 | 0.021 | 0.006 | 0.000 | 0.023 |
| Catalana | k NN | 0.808 | 0.734 | 0.685 | 0.727 | 0.772 | 0.582 | 0.305 | 0.758 |
| | WCI | 0.817 | 0.739 | 0.662 | 0.714 | 0.774 | 0.611 | 0.512 | 0.728 |
| | Clm | 0.189 | 0.273 | 0.178 | 0.239 | 0.030 | 0.010 | 0.014 | 0.042 |
| Duero | k NN | 0.800 | 0.824 | 0.743 | 0.818 | 0.514 | 0.623 | 0.594 | 0.507 |
| | WCI | 0.789 | 0.819 | 0.751 | 0.808 | 0.633 | 0.654 | 0.676 | 0.585 |
| | Clm | 0.319 | 0.449 | 0.151 | 0.372 | 0.020 | 0.010 | 0.007 | 0.011 |
| Ebro | k NN | 0.730 | 0.742 | 0.648 | 0.742 | 0.816 | 0.676 | 0.376 | 0.724 |
| | WCI | 0.744 | 0.745 | 0.674 | 0.744 | 0.798 | 0.808 | 0.517 | 0.775 |
| | Clm | 0.311 | 0.422 | 0.215 | 0.344 | 0.013 | 0.011 | 0.013 | 0.019 |
| Quadalquiver | k NN | 0.913 | 0.856 | 0.556 | 0.836 | 0.894 | 0.584 | 0.838 | 0.827 |
| | WCI | 0.916 | 0.863 | 0.701 | 0.841 | 0.940 | 0.661 | 0.951 | 0.909 |
| | Clm | 0.297 | 0.259 | 0.039 | 0.202 | 0.029 | 0.005 | 0.001 | 0.029 |
| Guadiana | k NN | 0.842 | 0.837 | 0.713 | 0.833 | 0.799 | 0.868 | 0.258 | 0.781 |
| | WCI | 0.842 | 0.852 | 0.889 | 0.857 | 0.821 | 0.825 | 0.378 | 0.812 |
| | Clm | 0.306 | 0.324 | 0.049 | 0.239 | 0.016 | 0.007 | 0.003 | 0.017 |
| Levante | k NN | 0.855 | 0.801 | 0.688 | 0.794 | 0.810 | 0.746 | 0.851 | 0.474 |
| | WCI | 0.873 | 0.798 | 0.690 | 0.805 | 0.916 | 0.788 | 0.948 | 0.604 |
| | Clm | 0.240 | 0.272 | 0.086 | 0.212 | 0.010 | 0.006 | 0.002 | 0.007 |
| Norte | k NN | 0.869 | 0.852 | 0.740 | 0.819 | 0.807 | 0.734 | 0.741 | 0.718 |
| | WCI | 0.870 | 0.844 | 0.751 | 0.806 | 0.842 | 0.754 | 0.835 | 0.751 |
| | Clm | 0.501 | 0.585 | 0.317 | 0.572 | 0.040 | 0.056 | 0.011 | 0.059 |
| Sur | k NN | 0.815 | 0.804 | 0.698 | 0.783 | 0.731 | 0.428 | 0.892 | 0.666 |
| | WCI | 0.840 | 0.823 | 0.716 | 0.790 | 0.813 | 0.640 | 0.964 | 0.720 |
| | Clm | 0.230 | 0.146 | 0.034 | 0.170 | 0.026 | 0.003 | 0.000 | 0.012 |
| Tajo | k NN | 0.884 | 0.849 | 0.725 | 0.874 | 0.982 | 0.525 | 0.922 | 0.394 |
| | WCI | 0.879 | 0.845 | 0.741 | 0.880 | 0.980 | 0.885 | 0.946 | 0.668 |
| | Clm | 0.298 | 0.367 | 0.104 | 0.280 | 0.008 | 0.006 | 0.004 | 0.011 |

scores among the different stations. However, the clustering method clearly outperforms the analog results in the case of extreme precipitation (Fig. 12b).

5. Conclusions and further remarks

We presented a new downscaling method for short-term forecast using a clustering technique: weighting clustering downscaling method. The method is computationally simpler and more efficient than the standard method of analogs. Validation results on 98 stations on the Iberian Peninsula show that best skill results (compared with the standard method of analogs) are obtained for extreme (or nonfrequent) events. This result indicates that the clustering process captures the periphery of the distribution of atmospheric patterns better than the fixed-size ensemble of nearest neighbors used in the standard analog method. We would like to mention that the results presented in this paper could be improved in the near future using the simulations of a new reanalysis project, ECMWF Re-Analysis-40 (ERA-40) covering the period from mid-1957 to 2001.

The algorithm proposed in this paper is not the first attempt to develop a clustering-based downscaling technique. For instance, the Classification and Regression Trees (CART) technique introduced by Hughes et al. (1993) and used later in Zorita et al. (1995) is also based on a clustering process of the atmospheric patterns (this is the first step of the method). However, the clustering algorithm used to define weather classes (clusters) is a binary decision tree, which works by splitting the values of the input variables (each of the grid values of the variables used to define the atmospheric patterns) so that a maximum separation of the precipitation occurrence distribution is attained in the resulting leaves of the tree. Each terminal node of the decision tree corresponds to a cluster (or weather class). This technique has been applied to problems of climatic change dealing with patterns with low complexity [sea level pressure (SLP), geopotential height at 500 mb]; these problems only require a few weather classes (clusters), and thus, the splitting process can be efficiently carried out considering individual input variables. However, an application of this technique in the short range would require

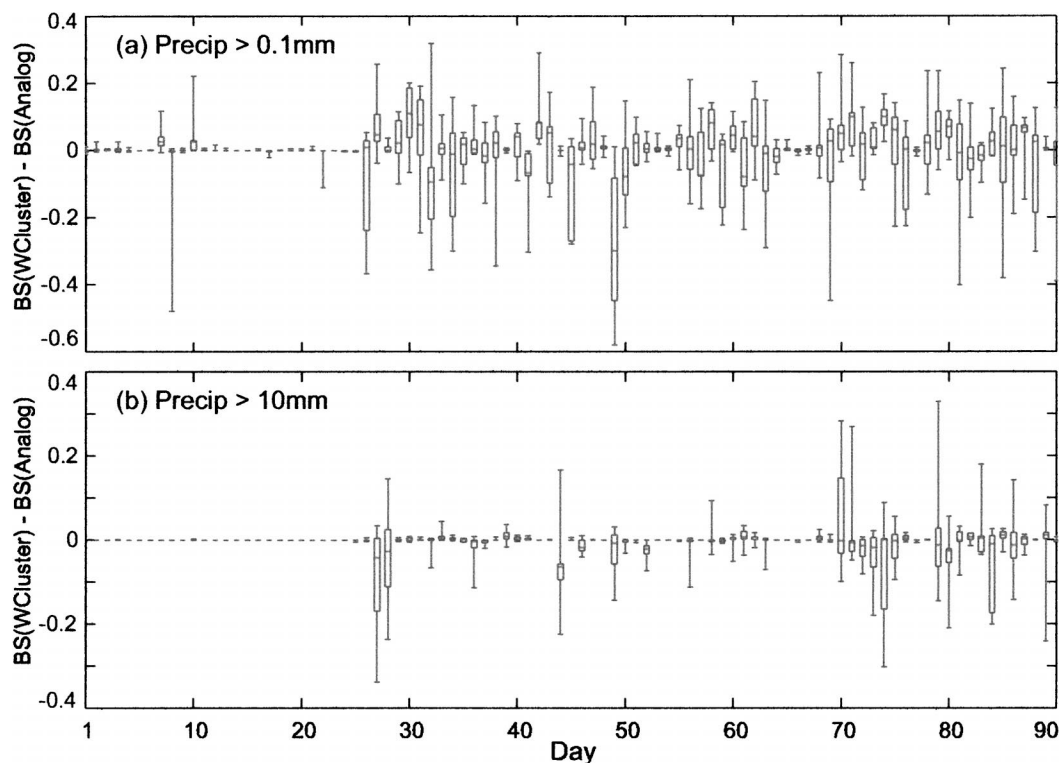


FIG. 12. Daily box-and-whisker plots of the precipitation BS differences between the weighted clustering (WCluster with $m = 400$ and $w = 4$) and standard analog (k -NN with $k = 50$) methods for a particular period of 90 days (SON 1998). For each day, the boxes represent the terciles for the 98 stations, and the whiskers show the extreme differences that occurred each day.

dealing with huge atmospheric patterns (for a general application) and a large number of clusters for which a clustering process based on a binary tree becomes inefficient.

Another clustering-based downscaling method was introduced by Cavazos (1997), who applied self-organizing maps (SOMS) to define weather classes and study the associated distributions of local precipitation. An SOM is a clustering technique developed in the field of neural networks with many interesting visualization capabilities. This technique has been recently shown to be a modification of the m -means algorithm with a topology preservation constraint (see, e.g., Hastie et al. 2001); this constraint alters the clustering process, decreasing the variance of the resulting clusters and reducing the skill of the clustering downscaling method proposed in this paper. Several comparison experiments corroborate this fact.

Finally, the term “short range” in the title of this paper refers to the fact that no use of ensemble forecast models is made. The skill of the downscaling method quickly decays by the fourth day, similar to the skill of the numerical model used as input. However, the method proposed in this paper is suitable to work with ensemble outputs, since each of the members of the ensemble will correspond to a particular cluster, and, hence, the whole ensemble will define a distribution on the clusters, ob-

taining a natural framework for working with probabilities. In this case, the visualization capabilities and the neighborhood preservation of the SOM can be fruitfully used. The application of the clustering method using SOMs to medium-range ensemble forecasts is in progress, and the results will be published in a separate paper.

Acknowledgments. The authors are grateful to the Instituto Nacional de Meteorología (INM) for providing us with partial support and the necessary data for this work. The authors are also grateful to the University of Cantabria, CSIC, and the Comisión Interministerial de Ciencia y Tecnología (CICYT; Grant REN2000-1572) for partial support of this work.

REFERENCES

- Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850.
- Bergman, M. J., and J. W. Delleur, 1985a: Kalman filter estimation and prediction of daily stream flows: 1. Review, algorithm, and simulation experiments. *Water Resour. Bull.*, **21**, 815–825.
- , and —, 1985b: Kalman filter estimation and prediction of daily stream flows: 2. Application to the Potomac River. *Water Resour. Bull.*, **21**, 827–832.
- Billet, J., M. Delisi, B. G. Smith, and C. Gates, 1997: Use of re-

- gression techniques to predict hail size and the probability of large hail. *Wea. Forecasting*, **12**, 154–164.
- Cavazos, T., 1997: Downscaling large-scale circulation to local winter rainfall in north-eastern Mexico. *Int. J. Climatol.*, **17**, 1069–1082.
- Duda, R. O., P. E. Hart, and D. G. Stork, 2000: *Pattern Classification*. 2d ed. Wiley-Interscience, 654 pp.
- Enke, W., and A. Spekat, 1997: Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Res.*, **8**, 195–207.
- Gardner, M. W., and S. R. Dorling, 1998: Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.*, **32**, 2627–2636.
- Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 533 pp.
- Hughes, J. P., D. P. Lettenmaier, and P. Guttorp, 1993: A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation. *Water Resour. Res.*, **29**, 3303–3316.
- Katz, R. W., and A. H. Murphy, Eds., 1997: *Economic Value of Weather and Climate Forecasts*. Cambridge University Press, 222 pp.
- Klein, W. H., and H. R. Glahn, 1974: Forecasting local weather by means of model output statistics. *Bull. Amer. Meteor. Soc.*, **55**, 1217–1227.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646.
- McGinnis, D. L., 1994: Predicting snowfall from synoptic circulation: A comparison of linear regression and neural networks. *Neural Nets: Applications in Geography*, B. Hewitson and R. G. Crane, Eds., Kluwer Academic, 79–99.
- Peña, J. M., J. A. Lozano, and P. Larrañaga, 1999: An empirical comparison of four initialization methods for the k-means algorithm. *Pattern Recognit. Lett.*, **20**, 1027–1040.
- Preisendorfer, R. W., and C. D. Mobley, 1988: *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, 425 pp.
- Toth, Z., 1991: Circulation patterns in phase space: A multinormal distribution? *Mon. Wea. Rev.*, **119**, 1501–1511.
- Van den Dool, H. M., 1989: A new look at weather forecasting through analogues. *Mon. Wea. Rev.*, **117**, 2230–2247.
- von Storch, H., 1999: On the use of “inflation” in statistical downscaling. *J. Climate*, **12**, 3505–3506.
- Wilby, R. L., and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr.*, **21**, 530–548.
- Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489.
- , J. P. Hughes, D. P. Lettenmaier, and H. von Storch, 1995: Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. *J. Climate*, **8**, 1023–1042.

Copyright of Monthly Weather Review is the property of American Meteorological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.